



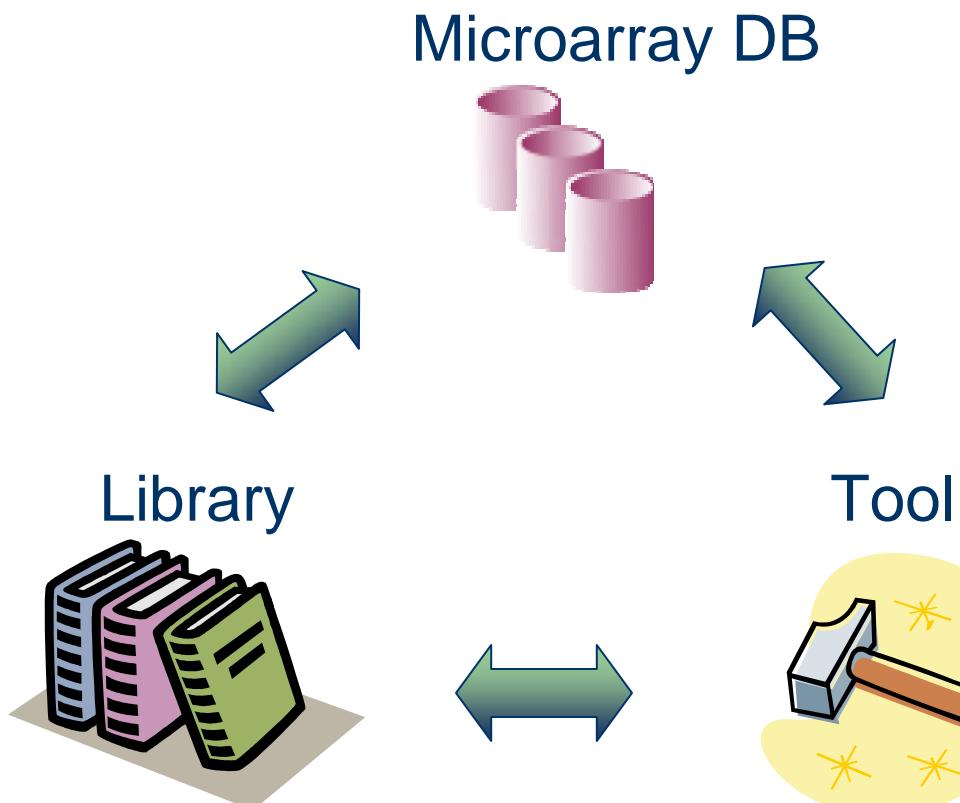
ArrayTrack_{3.4.0} Demonstration

National Center for Toxicological Research
U.S. Food and Drug Administration
3900 NCTR Road,
Jefferson, AR 72079





ArrayTrack Overview



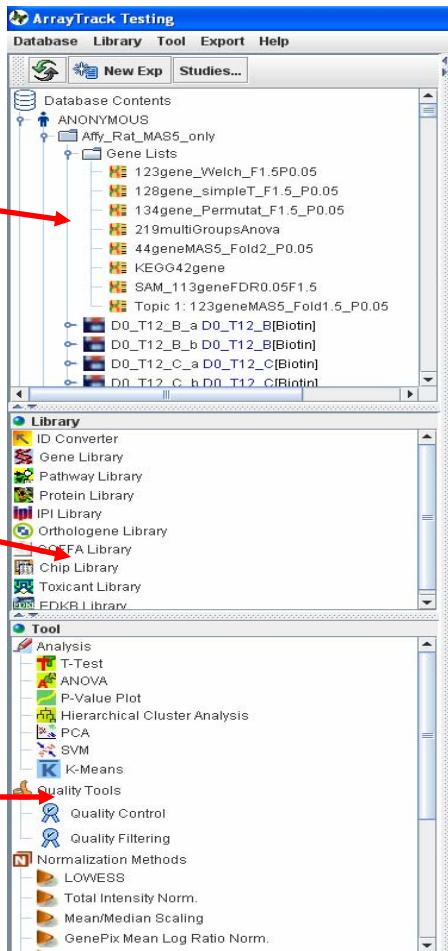
An integrated environment for microarray data management, analysis and interpretation.

Using ArrayTrack, the user can apply analysis method from Tool to microarray DB and then get information from the linked LIB for biological interpretation



ArrayTrack Overview

Microarray DB



Libraries

ID Converter

Gene Library

Pathway Library

Protein Library

IPI Library

Orthologene Library

L00SFA Library

Chip Library

Toxicant Library

EDKRLI Library

Tools

Analysis

T-Test

ANOVA

P-Value Plot

Hierarchical Cluster Analysis

PCA

SVM

K-Means

Quality Tools

Quality Control

Quality Filtering

Normalization Methods

LOWESS

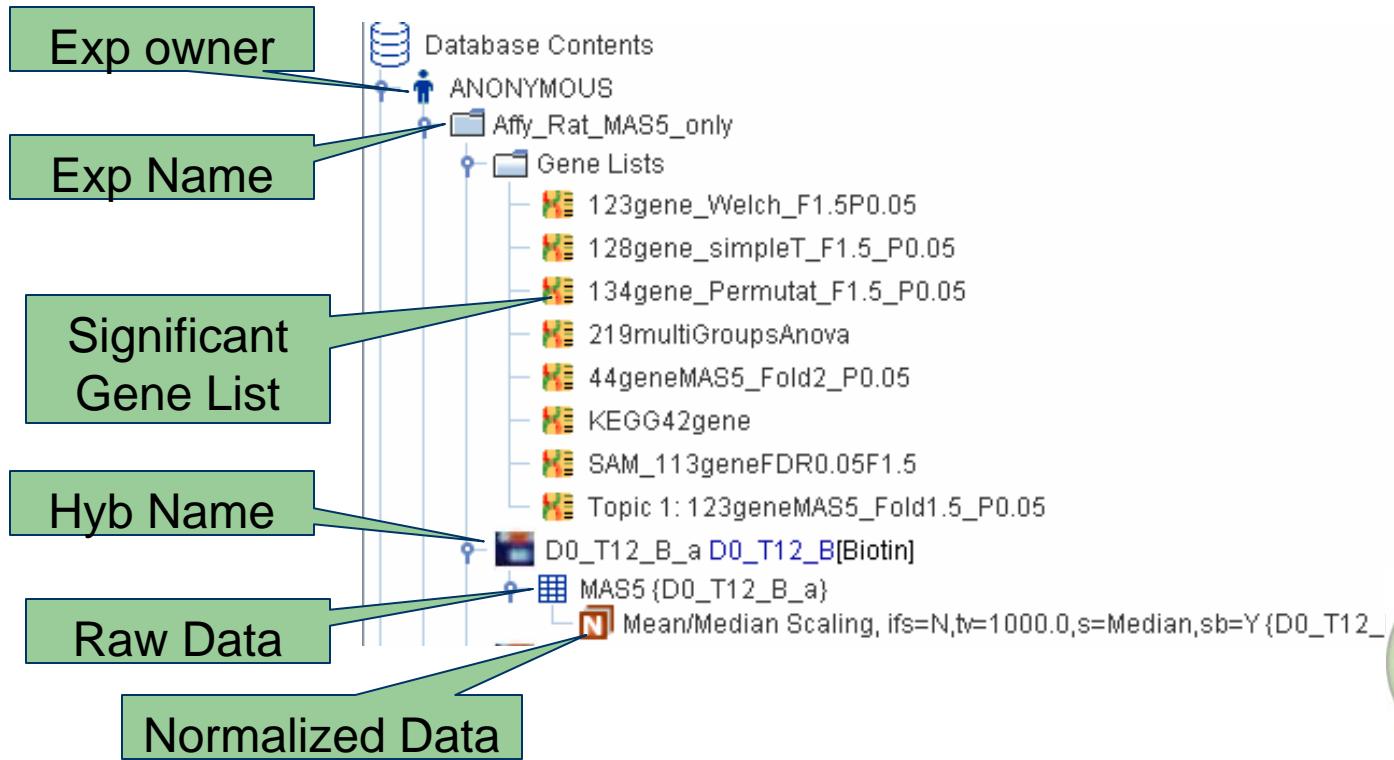
Total Intensity Norm.

Mean/Median Scaling

GenePix Mean Log Ratio Norm.

Microarray DB

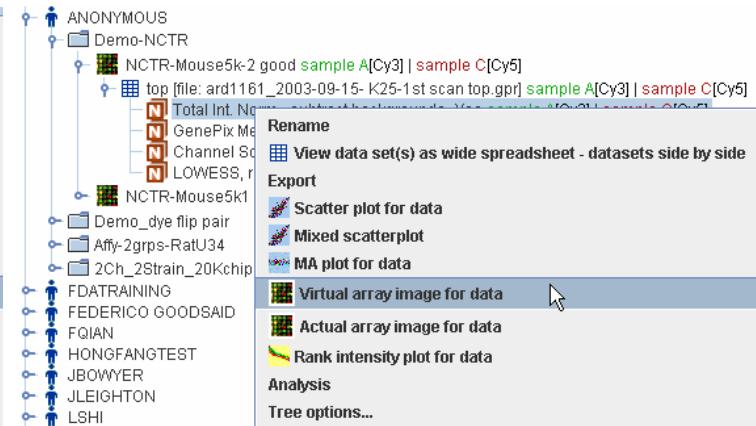
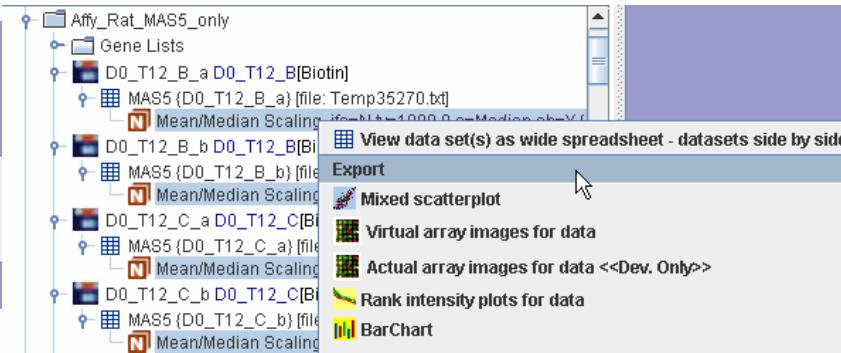
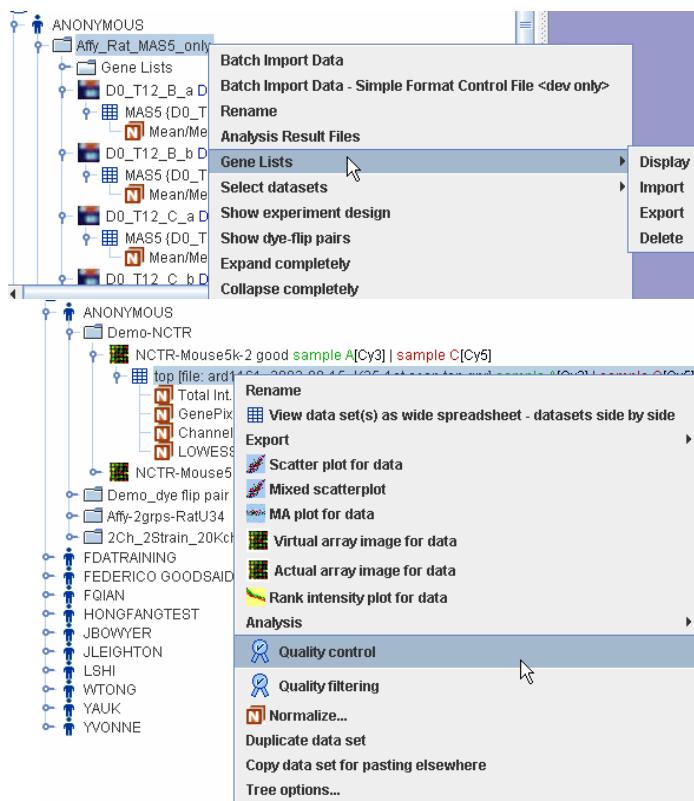
Data is organized as a hierarchical tree structure:





Exploring DB

- Right-click exp, hybridization, raw data and normalized data to access various functions



Exploring DB

Experiment design

The screenshot shows a software application window titled "Hybridizations". The menu bar includes "Edit", "View", and "Input/Output". Below the menu is a toolbar with icons for saving, opening, and closing. A "Filter->" button is also present.

The main area displays a table with the following columns: Hybridization, SAMPLE 1, LABEL 1, ARRAYTYPENAME, CHANNEL, SPECIES 1, ASSAY 1, and CELLTYPE 1. The table contains the following data:

	Hybridization	SAMPLE 1	LABEL 1	ARRAYTYPENAME	CHANNEL	SPECIES 1	ASSAY 1	CELLTYPE 1
1	D0_T12_B_a	D0_T12_B	Biotin	Affy_RT-U34	One Channel	Rat	In Vitro	Hepatocytes
2	D0_T12_B_b	D0_T12_B	Biotin	Affy_RT-U34	One Channel	Rat	In Vitro	Hepatocytes
3	D0_T12_C_a	D0_T12_C	Biotin	Affy_RT-U34	One Channel	Rat	In Vitro	Hepatocytes
				Affy_RT-U34	One Channel	Rat	In Vitro	Hepatocytes
				Affy_RT-U34	One Channel	Rat	In Vitro	Hepatocytes
				Affy_RT-U34	One Channel	Rat	In Vitro	Hepatocytes
				Affy_RT-U34	One Channel	Rat	In Vitro	Hepatocytes
				Affy_RT-U34	One Channel	Rat	In Vitro	Hepatocytes
				Affy_RT-U34	One Channel	Rat	In Vitro	Hepatocytes
				Affy_RT-U34	One Channel	Rat	In Vitro	Hepatocytes
				Affy_RT-U34	One Channel	Rat	In Vitro	Hepatocytes
				Affy_RT-U34	One Channel	Rat	In Vitro	Hepatocytes
				Affy_RT-U34	One Channel	Rat	In Vitro	Hepatocytes
				Affy_RT-U34	One Channel	Rat	In Vitro	Hepatocytes
				Affy_RT-U34	One Channel	Rat	In Vitro	Hepatocytes
				Affy_RT-U34	One Channel	Rat	In Vitro	Hepatocytes

Below the table, a context menu is open with the following options:

- Batch Import Data
- Batch Import Data - Simple Format Control File <dev only>
- Rename
- Analysis Result Files
- Gene Lists
- Select datasets
- Show experiment design
- Show dye-flip pairs
- Expand completely
- Collapse completely



Searching Libraries



There are nine libraries in ArrayTrack.
All the libraries are interlinked.

The libraries integrate gene, protein, pathway and other data allowing data interrogation and mining of data across data types.

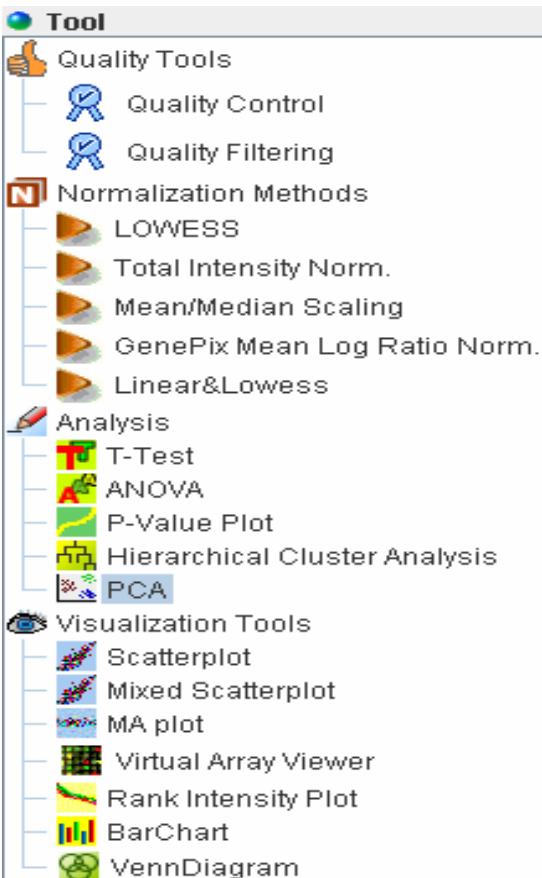
Several ways to activate these libraries

- From library panel
- From the Library pull-down menu
- From the results of analysis (e.g. T-test)





Analysis Tools



Tools:

• Quality Tools

Provides various visual plots and numerical parameters for measuring the quality of a hyb, and filtering the unwanted spots.

• Normalization Methods

Correct systematic variations in microarray data introduced by experimental factors

• Analysis

Provides different methods to evaluate the microarray data

• Visualization Tools

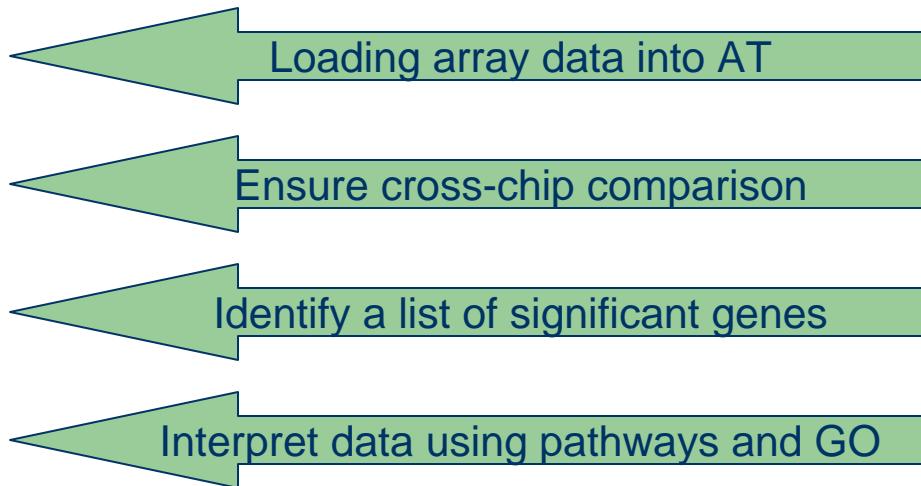
Provides a direct view to identify abnormalities within data.





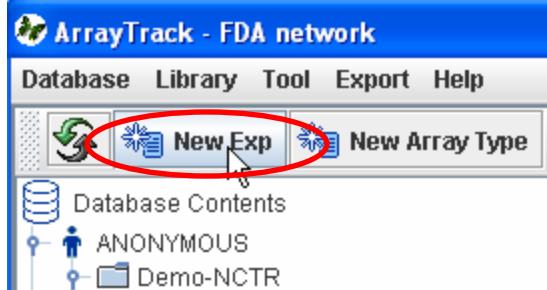
Microarray Data Processing

- Importing data
- Normalization
- Gene Selection
- Interpretation





Importing Data



1. Create experiment
2. Create array type (if not existed in AT Chip LIB)
3. Use batch import wizard (see detail in Tutorial 9)

A screenshot of the "Input Form" window. It contains two main tabs: "Experiment Design" and "Hybridization and Data".
Experiment Design Tab:

- Experiment ID: Demo-NCTR
- Experimenter: anonymous
- Institute: NCTR
- Exp Types: Temperature shock; normal vs baseline
- Key words: rat, liver, circadian
- Exp Description: Ad lib fed rats at 6 hr. intervals; examining liver changes
- Phenotype Anchoring: Ad lib fed rats at 6 hr. intervals; examining liver changes
- Comments:

Buttons: Clean Form, Print Form, Help, Save Exp, Import, Export.
Hybridization and Data Tab:

- Hybridization: Select
- Sample ID: Select
- Label: Select
- Notes: OC Notes for Label, OC Notes for Hyb

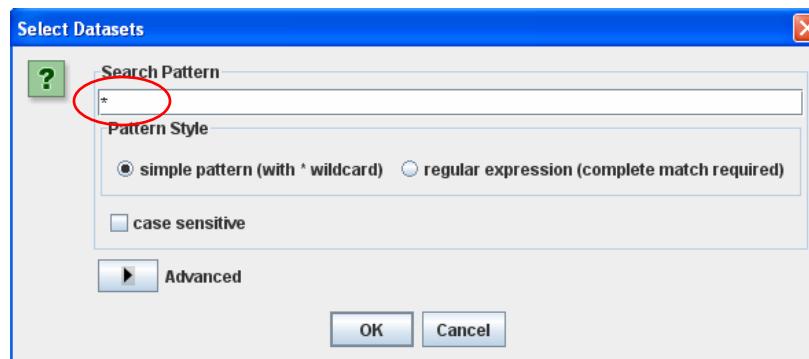
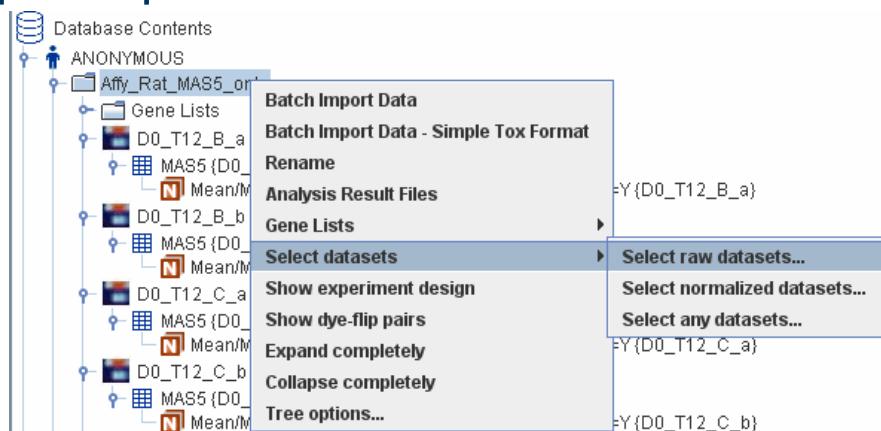
Data Import: Import

Normalization

- To remove systematic variations across chips and ensure a valid cross-chip comparison

1. Right-click an experiment
2. Choose “Select datasets”
->“Select raw dataset...”

3. Type wildcard * if you want to select all the data.
Click OK button.





Normalization (-continued)

3. Right-click any highlighted raw data
4. Select “Normalize...”

The screenshot shows the Array Track software interface. On the left is a tree view of database contents under an anonymous user. The tree includes nodes for raw data sets (Affy_Rat_MAS5_only, D0_T12_B_a, D0_T12_B_b, D0_T12_C_a, D0_T12_C_b, D0_T12_D_a, D0_T12_D_b) and processed data sets (MAS5 (D0_T12_B_a), MAS5 (D0_T12_B_b), MAS5 (D0_T12_C_a), MAS5 (D0_T12_C_b), MAS5 (D0_T12_D_a), MAS5 (D0_T12_D_b)). A context menu is open over the first MAS5 node, listing various analysis options. The "Normalize..." option is at the bottom of the menu, highlighted with a blue background.

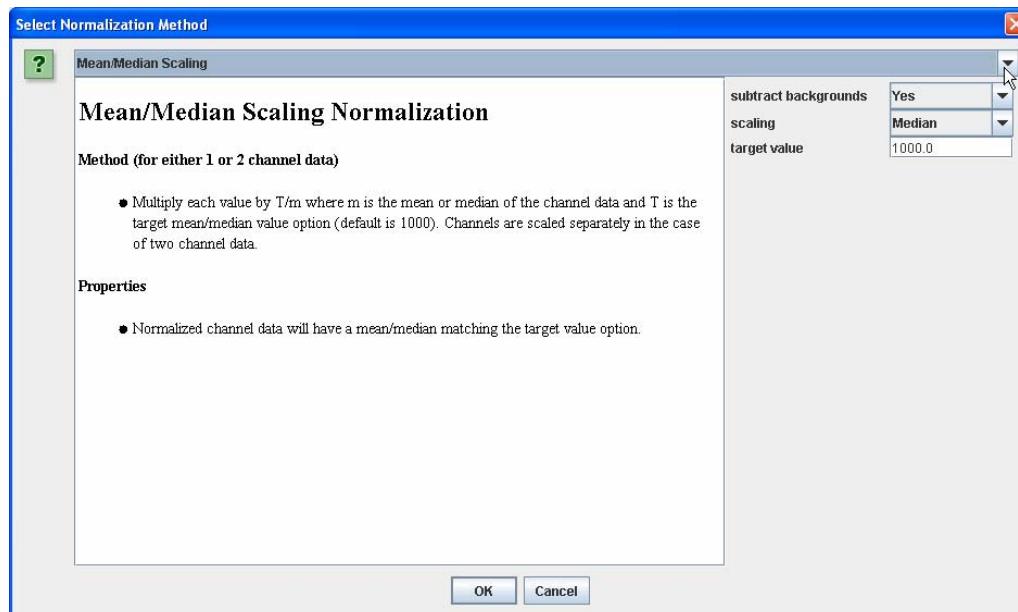
- View data set(s) as wide spreadsheet - datasets side by side
- Export
- Convert affy cel files to probe sets
- Mixed scatterplot
- Virtual array images for data
- Actual array images for data <<Dev. Only>>
- Rank intensity plots for data
- BarChart
- Create gene list by data filtering...
- Analysis
- Quality Control
- Normalize...**





Normalization (-continued)

- Choose normalization method



- For Affy data, choose “Mean/Median Scaling”
- For 2-channel data, the default method is “Lowess”



Gene Selections

Determining a list of genes that are differentially expressed between control and treatment.

Two types of experiment:

Single testing: \longrightarrow 1 gene $P < 0.05$ low error rate

Multiple testing: \longrightarrow n genes $P = nP_i$ If $P_i = 0.05$, high error rate

e.g., If $n=10$ and $P_i=0.05$, $P=0.5$ for family-wise error

Error rate for the exp

Select a gene list based on:

P value



Bonferroni correction
 P_i/n

Low sensitivity

Low power



False discovery rate (e.g., Benjamini & Hochberg, p-value plot)

Permutation t-test (e.g., SAM)

Volcano plot (combination of p and fold change)



Gene Selection

- Highlight and Right-click the experiment.
- Select “raw datasets...” and click “OK”
- Right-click the highlighted raw data
- Choose “Analysis->T-test”

The screenshot shows the Array Track software interface. On the left, there is a tree view of experiments and datasets. A right-click context menu is open over a dataset named 'D0_T12_B_a'. The menu has several options: 'View data set(s) as wide spreadsheet - datasets side by side', 'Export', 'Convert affy cel files to probe sets', 'Mixed scatterplot', 'Virtual array images for data', 'Actual array images for data <>Dev. Only>', 'Rank intensity plots for data', 'BarChart', 'Create gene list by data filtering...', 'Analysis' (which is currently selected), 'Quality Control', 'Normalize...', 'Duplicate data sets', 'Copy data sets for pasting elsewhere', 'Studies', and 'Tree options...'. The 'Analysis' section is expanded, showing sub-options: 'T-Test/ANOVA' (which is selected), 'SAM-Test', 'R-Interface', 'Correlation Matrix', 'Correlate Study Data vs. Gene Expression Data', 'T-Test with custom data options', 'ANOVA with custom data options', 'Hierarchical Cluster Analysis', and 'Principal Component Analysis'. In the bottom right corner of the interface, there is a small graphic of a butterfly with a grid pattern on its wings.



Gene Selection (-continued)

Select Dataset Group Assignments for T-Test / ANOVA / Pairwise Tests

Assign Data Sets Into Groups

	Hybridization	SAMPLE 1	Datafile	LABEL 1	ARRAYTYPE	NAME	CHANNEL	SPECIES	ASSAY 1	F
1 (1)	D0_T12_B_a	D0_T12_B	B12Cd000A.CEL	Temp3962.t	Biotin	Affy_RT-U34	One Channel	Rat	In Vitro	N
2 (1)	D0_T12_B_b	D0_T12_B	B12Cd000B.CEL	Temp3964.t	Biotin	Affy_RT-U34	One Channel	Rat	In Vitro	N
3 (1)	D0_T12_C_a	D0_T12_C	C12Cd000A.CEL	Temp3970.t	Biotin	Affy_RT-U34	One Channel	Rat	In Vitro	N
4 (1)	D0_T12_C_b	D0_T12_C	C12Cd000B.CEL	Temp3972.t	Biotin	Affy_RT-U34	One Channel	Rat	In Vitro	N
5 (1)	D0_T12_D_a	D0_T12_D	D12Cd000A.CEL	Temp3978.t	Biotin	Affy_RT-U34	One Channel	Rat	In Vitro	N
6 (1)	D0_T12_D_b	D0_T12_D	D12Cd000B.CEL	Temp3980.t	Biotin	Affy_RT-U34	One Channel	Rat	In Vitro	N
7 (2)	D2_T12_B_a	D2_T12_B	B12Cd2.00A.CEL	Temp3966.t	Biotin	Affy_RT-U34	One Channel	Rat	In Vitro	N
8 (2)	D2_T12_B_b	D2_T12_B	B12Cd2.00B.CEL	Temp3968.t	Biotin	Affy_RT-U34	One Channel	Rat	In Vitro	N
9 (2)	D2_T12_C_a	D2_T12_C	C12Cd2.00A.CEL	Temp3974.t	Biotin	Affy_RT-U34	One Channel	Rat	In Vitro	N
10 (2)	D2_T12_C_b	D2_T12_C	C12Cd2.00B.CEL	Temp3976.t	Biotin	Affy_RT-U34	One Channel	Rat	In Vitro	N
11 (2)	D2_T12_D_a	D2_T12_D	D12Cd2.00A.CEL	Temp3982.t	Biotin	Affy_RT-U34	One Channel	Rat	In Vitro	N
12 (2)	D2_T12_D_b	D2_T12_D	D12Cd2.00B.CEL	Temp3984.t	Biotin	Affy_RT-U34	One Channel	Rat	In Vitro	N

2 groups, sizes = [6, 6]

T-Test fold changes are computed as grp 1/grp 2, so "up" regulation will mean grp 1 > grp 2 in any further analysis.

Next >

Assign the data into 2 groups
-different dose
-different time
-or different animal

Note: always put control in group 2



Gene Selection (-continued)

Select Dataset Group Assignments for T-Test / ANOVA / Pairwise Tests

Test Type (Consistent with group selections)

T-Test

T-Test Options

P values from dist.: Welch t-test Simple t-test One class vs. mean:

P values from permutations: All Limit to:

Filtering with a gene list

Only include genes from gene list

Gene identifiers to include

Genbank Acc Gene Mfr ID LOCUSID UNIGENEID GENENAME
 CLONEID GEN_DESCR_MFR REFSEQ SPOTID

Dataset Naming
Hybridization names are always included.
 add sample name(s) to hybridization names
 add dye name(s) to hybridization names

Data options

Subtract backgrounds when present (raw datasets only)
 Apply log (base 2) to expression values
 Exclude spots flagged as bad



Gene Selection (-continued)

T-Test Results

File Selected-Spot All-Spots Advanced

	Genbank Acc	Gene Mfr ID	LOCUSID	GENENAME	REFSEQ	SPOTID	▲ (1) P	Abs Fold C...	Fold Chang...
1	L26267	L26267_at	81736	Nfkb1		516822	0	2.2286	2.2286
2	L15079	L15079mR...	24891	Abcb4	NM_012690	516807	0	1.8323	1.8323
3	D87336	D87336_g...	287552	Blmh		516718	0.0001	1.7665	1.7665
4	X07365	X07365_s...	24404	Gpx1		517129	0.0001		
5	M55534	M55534m...	25420	Cryab	NM_012935	516911	0.0001		
6	J02722	J02722cds...	24451	Hmxo1		516748	0.0001		
7	E00778	E00778cds...	24296	Cyp1a1		516730	0.0001		
8	AA848563	AA848563_s...		Hspa1a	/// ...	516506	0.0001		
9	AA818604	rc_AA8186...		Hspa1a	/// ...	517260	0.0001		
10	Z75029	Z75029_s...	294254	Hspa1b		517240	0.0004	7.39	7.39
11	U01344	U01344_g...	116631	Nat1	NM_053853	516979	0.0005		
12	M99169	M99169_at	81771	Rps6ka1	NM_031107	516964	0.0005		1.56
13	AA108277	AA108277...	288444	Hspf1		516484	0.0007	4.1286	0.2422
14	AI229655	rc_AI229655...	363249	Ctdsp1		516484	0.0007	1.5386	1.5386
15	D16478	D16478_g...	170670	Hadha	NM_139078	517416	0.0008	1.5541	1.5541
16	AI103396	rc_AI10339...				517415	0.0011	2.3716	2.3716
17	E00717	E00717_IT	24298	Cyn1		516729	0.0011	2.7807	2.7807

123 genes

Significance Filtering

P Values < 0.05 without adjustment

Target False Discovery Rate (FDR):

Select # genes by lowest p-values

Mean Channel Intensities > Bad Flags <=

Abs Fold Change > 1.5 Advanced>

Apply Filters Clear Filters

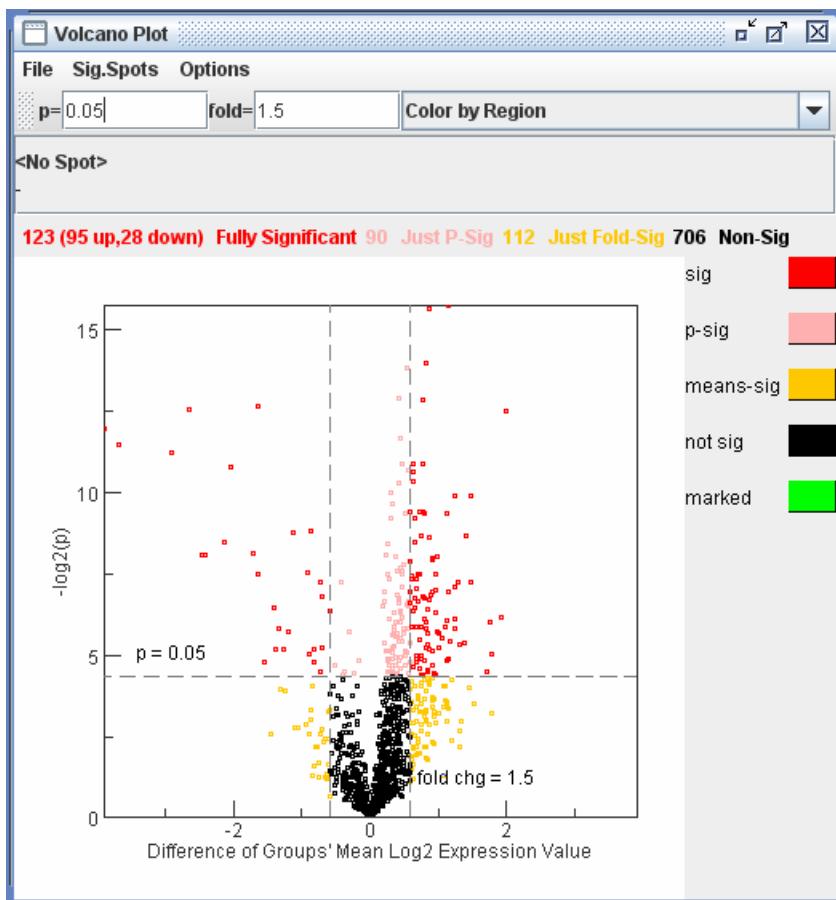
P-Value Plot Create Sig. Gene List HCA PCA K-Means Volcano Plot

Parameters used to filter the genes to get significant gene list

Access other analysis methods like HCA, PCA, Volcano plot, etc.



Gene Selection (-continued)



Volcano Plot



Significant Gene List

- Create
- Display
- Import
- Export
- Delete

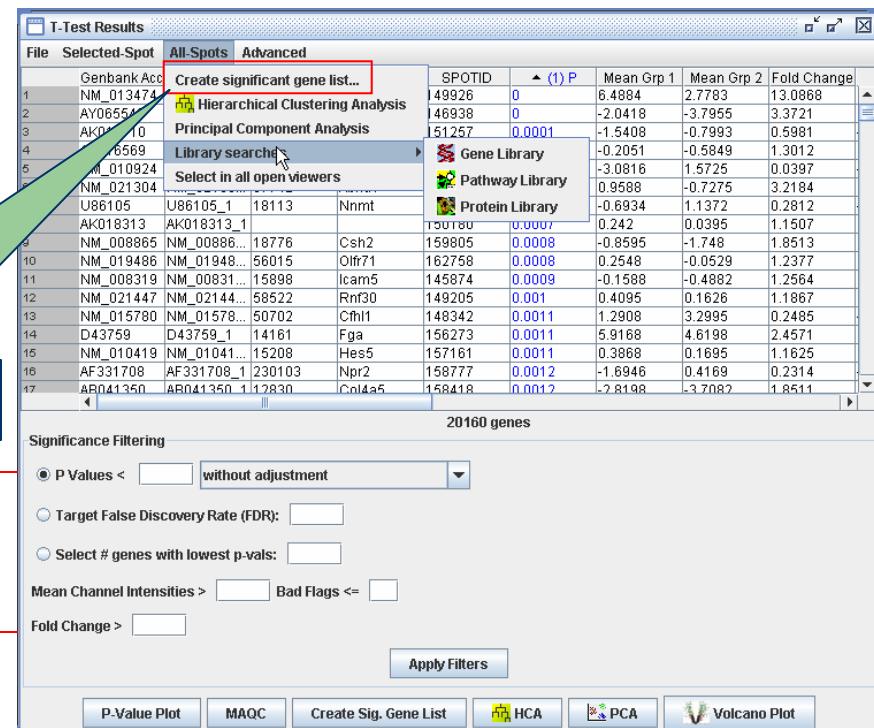
Significant Gene List(-cont)

- Create significant gene list

From T-test/Anova result
set filter criteria
e.g. P-value, fold change
to get significant gene list

Create significant gene list

Filter criteria



Significant Gene List(-cont)

- **Display/Import/Export/Delete Significant Gene List**

Right-click the experiment name

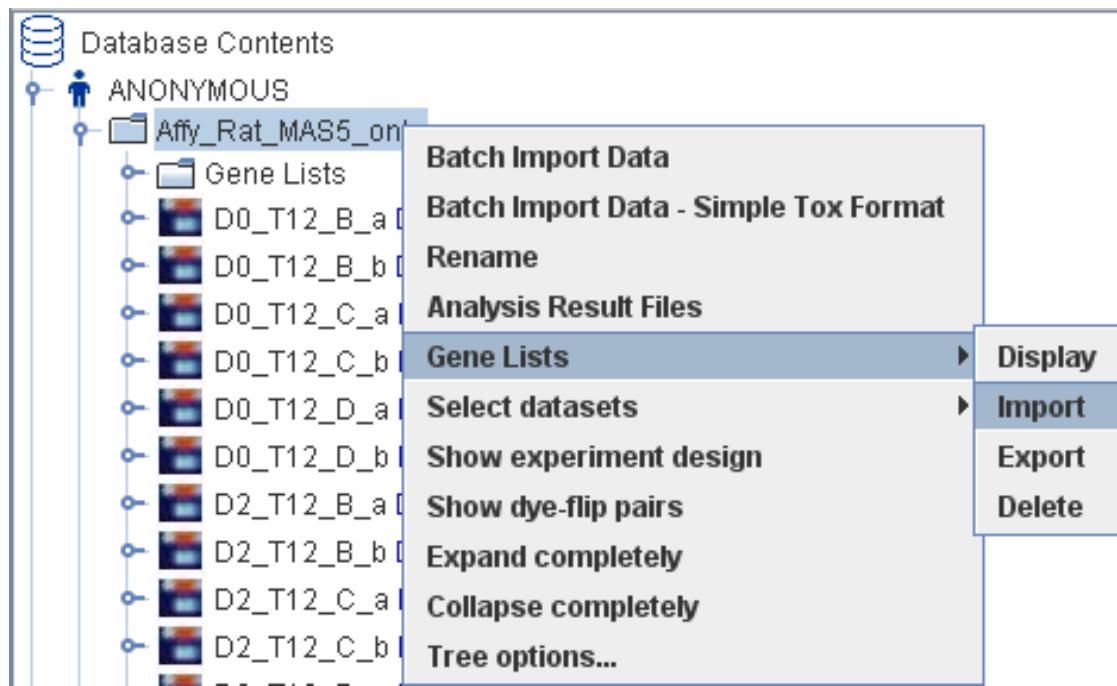
select Significant Gene Lists

select Display

Import

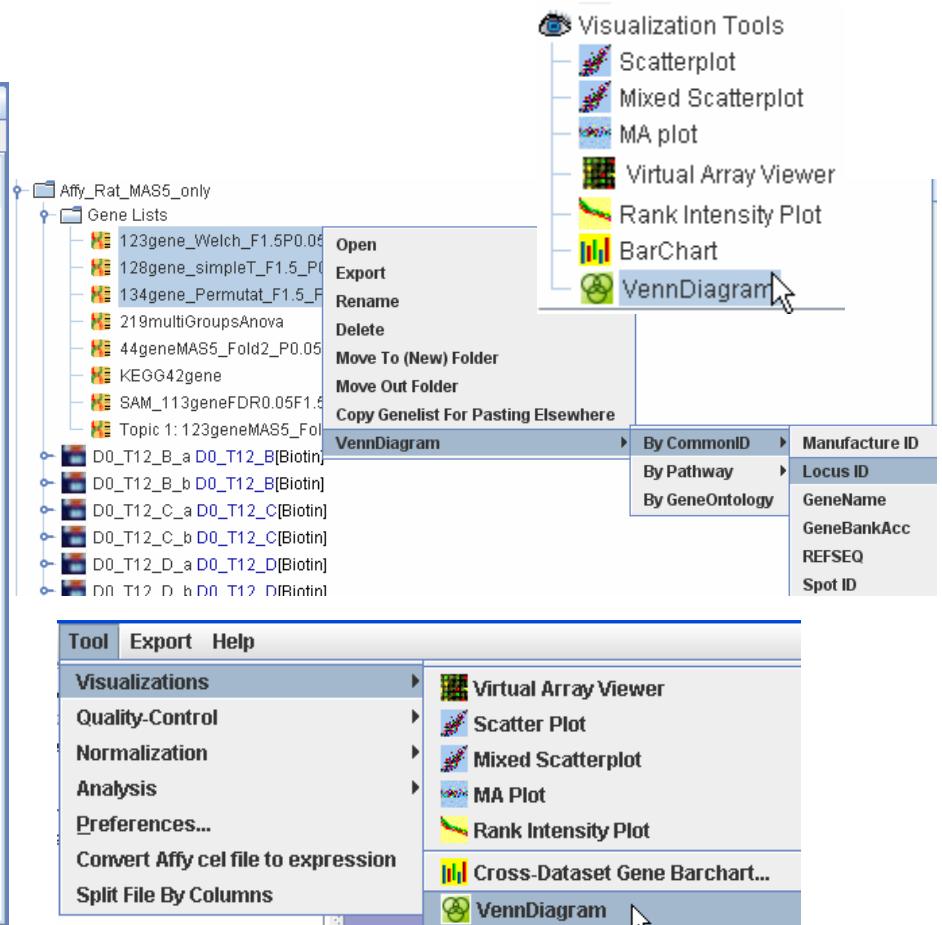
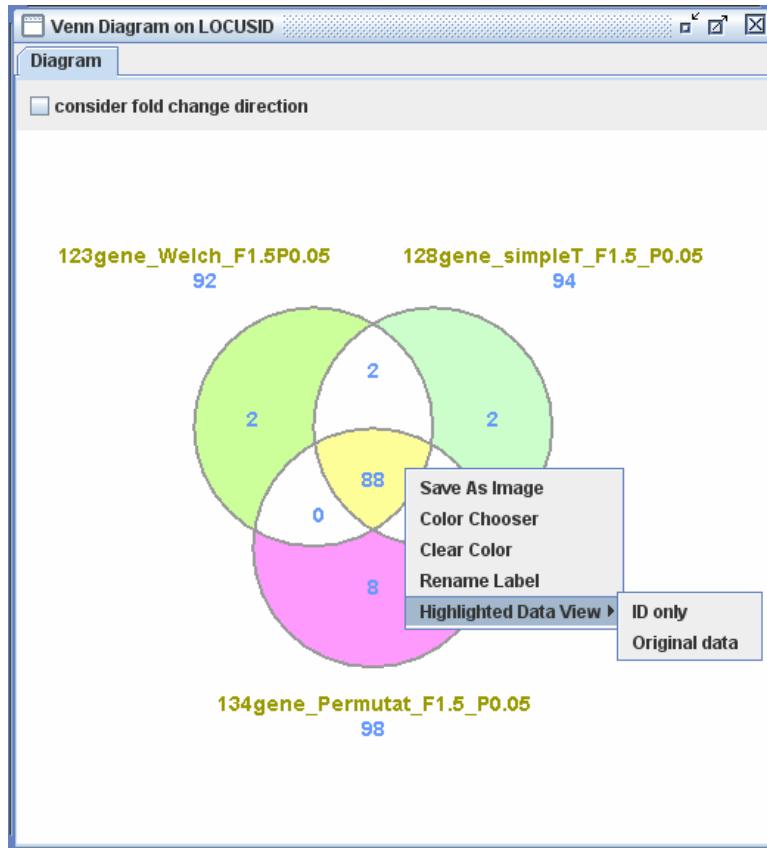
Export

Delete



Significant Gene List(-cont)

- **VennDiagram** - get common genes from 2~3 significant gene, protein, or metab lists.





Interpretation

Link the significant genes to various libraries for data interpretation

The screenshot shows the SIGNIFICANT_GENELIST software interface. On the left, there's a tree view of 'Database Contents' under 'ANONYMOUS' and a 'Library' section with options like ID Converter, Gene Library, Pathway Library, Protein Library, IPI Library, Orthologene Library, GOFFA Library, and Chip Library. The main window displays a table titled 'SIGNIFICANT_GENELIST' with columns: GENELIST_NAME *, EXPID *, GENEBOOKACC, GENENAME, LOCUSID, FOLD, and PVALUE. A red oval highlights the 'Chip Lib' button in the toolbar above the table. Another red oval highlights the 'GOFFA' and 'Orthologene' buttons in the same toolbar. The table contains 14 rows of gene information, such as Cyp1a1, Gpx1, Hmox1, Hspb1, etc., with their respective EXPID, GENEBOOKACC, LOCUSID, FOLD change, and P-value.

GENELIST_NAME *	EXPID *	GENEBOOKACC	GENENAME	LOCUSID	FOLD	PVALUE
123gene_Welch_F1.5P0.05	650	E00717	Cyp1a1	24296	2.7807	0.0011
128gene_simpleT_F1.5	650	E00778	Cyp1a1	24296	3.9988	0.0002
134gene_Permutat_F1.5_P0.	650	X07365	Gpx1	24404	1.7031	0.0001
44geneMASS5_Fold2_P0.05	650	X12367	Gpx1	24404	1.716	0.0015
SAM_test1	650	J02722	Hmox1	24451	0.16	0.0002
Table 1: 123genesMASS5_Fold4	650	M86389	Hspb1	24471	0.6191	0.0278
	650	AA998683	Hspb1	24471	0.6025	0.0068
	650	AI176658	Hspb1	24471	0.6102	0.0092
	650	L16764	Hspa1a // Hspa1b_mappe	24472	0.3038	0.0037
	650	U65007	Met	24553	2.195	0.0016
	650	X56420	Mpg	24561	2.1997	0.0348
	650	X96394	Abcc1	24565	1.7617	0.0277
	650	M11794	Mt1a	24567	0.5328	0.0056
	650	X68394	Nras	24605	1.5958	0.0178



Interpretation (continued)

Gene Library

Specify ID Type:

- GenBankAcc
- UnigeneID
- LocusID
- SwissProtAcc
- IMAGEID
- GEN_ID_MFR
- GeneSymbol
- Hs Mm Rn

Enter Searching Data:

Search within result

Message:
unique search ID number :94

GENENAME	DESCRIPTION	SPECIES	CHROMLOC	LOCUSID	BIOLOGI
Nat1	N-acetyltransferase 1 (arylamine N-acetyltransferase)	Rattus norvegicus	16p14	116631	metabolism
Tap1	transporter 1, ATP-binding cassette, subfamily A (ABCA) member 1	Rattus norvegicus	20p12	24811	defense response;immune response
Merf2d	myocyte enhancer factor 2D	Rattus norvegicus	2q34	81518	
Adprt	ADP-ribosyltransferase 1	Rattus norvegicus	13q26	25591	protein amino acid ADP-ribosylation
Dpyd	dihydroxyimidine dehydrogenase	Rattus norvegicus	2q41	81656	pyrimidine base catabolism
Mapk7	mitogen-activated protein kinase 7	Rattus norvegicus	10q23	114509	
LOC288591	similar to transmembrane protein inducible by retinoic acid	Rattus norvegicus	12q12	288591	
Dusp6	dual specificity phosphatase 6	Rattus norvegicus	7q13	116663	cell differentiation;protein amino acid phosphorylation
LOC288444	similar to heat shock protein 105 kDa	Rattus norvegicus	12p11	288444	
Gstm5	glutathione S-transferase, mu 5	Rattus norvegicus	2q34	64352	
Nras	neuroblastoma RAS viral (v-ras) oncogene homolog	Rattus norvegicus	2q34	24605	Ras protein signal transduction;cell cycle
Rara	retinoic acid receptor, alpha	Rattus norvegicus	10q31	24705	regulation of anti-apoptosis;regulation of gene expression
Dia1	diaphorase 1	Rattus norvegicus	7q34	25035	electron transport
Cd44	CD44 antigen	Rattus norvegicus	3q31	25406	cell adhesion;defense response
Nat2	N-Acetyltransferase-2	Rattus norvegicus	16p14	116632	metabolism
LOC308047	similar to heat shock protein, DNAJ-like	Rattus norvegicus	1p11	308047	
Met	met proto-oncogene	Rattus norvegicus	4q21	24553	adult behavior;brain development
Hspb1	heat shock 27kDa protein 1	Rattus norvegicus	12q12	24471	
Mapk14	mitogen activated protein kinase 14	Rattus norvegicus	20p12	81649	angiogenesis;protein amino acid phosphorylation
LOC363249	similar to golli-interacting protein	Rattus norvegicus	9q33	363249	
Hadha	hydroxyacyl-Coenzyme A dehydrogenase	Rattus norvegicus	6q12	170670	fatty acid metabolism;metabolism
Camlg	calcium modulating ligand	Rattus norvegicus	17p14	81715	
Cdk7	cyclin-dependent kinase 7 (MO15 homolog)	Rattus norvegicus	2q12	171150	cell cycle;cell cycle;cytokinesis;microtubule organization
Cox6a2	cytochrome c oxidase, subunit VIa, polypeptide 2	Rattus norvegicus	1q36	25278	electron transport
Mit1a	Metallothionein	Rattus norvegicus	18p12	24567	nitric oxide mediated signal transduction
Hspa1b	heat shock 70kD protein 1B	Rattus norvegicus	20p12	294254	response to heat
LOC293991	similar to NADH dehydrogenase (ubiquinol) reductase subunit 1	Rattus norvegicus	1q54	293991	
LOC315994	similar to Expressed sequence AI8746	Rattus norvegicus	8q32	315994	
Map3k1	mitogen activated protein kinase kinase 1	Rattus norvegicus	2q14	116667	protein amino acid phosphorylation

- The significant genes are listed here in Gene Library.
- Can search and sort the Gene library
- There are links to other Libraries(Kegg, Pathart)





Interpretation (continued)

- KEGG – Kyoto Encyclopedia of Genes and Genomes
<http://www.genome.jp/kegg/>
- KEGG is a suite of databases and associated software.
- KEGG Pathway database provides the information of metabolic, regulatory and disease pathways; Most of them are metabolic pathways.





Interpretation (continued)

PathArt (Jubilant) – a pathway database

- The Pathways (over 600 mammalian disease and signaling)
- The Pathways is a collection of manually curated information from literature and public domain databases.

In ArrayTrack

	Human	Rat	Mouse
Kegg	200	187	193
PathArt	587	151	297



Interpretation (continued) Kegg

KEGG Pathway Analysis results:

Gene	Map	Category	Fisher P Value
Gstm5	Glutathione metabolism(mmu00480)	Metabolism of Other Amino Acids/Metab...	0.00370292
Gpx1			
Ptgs2	Prostaglandin and leukotriene metaboli...	Lipid Metabolism/Metabolic pathway	0.25805128
Hmox1	Porphyrin and chlorophyll metabolism(m...	Metabolism of Cofactors and Vitamins/M...	0.13404114
Dusp6			
Hspb1			
Mapk7			
Mapk14			
Tnfrsf1a			
Tgfb2			
Nras			
Mapk9			
Map3k1			
Met			
Ngfr			
Tgfb2			
Tnfrsf1a			
Tnfrsf1a	Cytokine-cytokine receptor interaction(m...	Regulatory pathway	0.17855902
Apoptosis(mmu04210)		Regulatory pathway	0.58437738
Mapk9	Wnt signaling pathway(mmu04310)	Regulatory pathway	0.77999365
Tgfb2	TGF-beta signalling pathway(mmu04350)	Regulatory pathway	0.55443832
Mapk14	Toll-like receptor signalling pathway(mmm...	Regulatory pathway	0.23520864
Mapk9			
Csnk1d	Circadian rhythm(mmu04710)	Regulatory pathway	0.11721566
Mapk14	Parkinson's disease(mmu05020)	Regulatory pathway	0.03611750
Mapk9			
Gpx1	Amyotrophic lateral sclerosis (ALS)(mm...	Regulatory pathway	0.15056025

Input genes = 60, 22 genes found, 38 not found, Total 23 pathway maps.

Annotations:

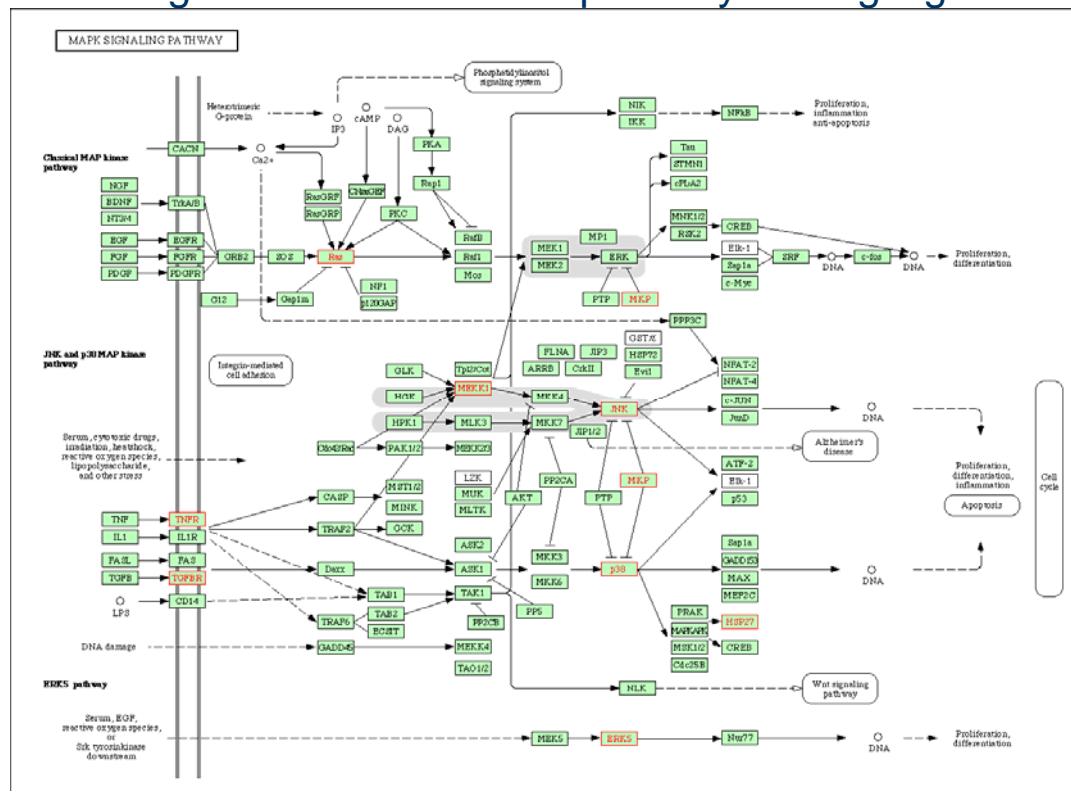
- Genes involved in a pathway
- Pathway name
- Pathway category
- Statistical significance of the pathway

Kegg Pathway



Interpretation (continued) Kegg

Double-click a specific pathway, the pathway map will be displayed and the genes involved in the pathway are highlighted.



Interpretation (continued) PathArt

Genes Pathways Physiology/disease Statistical significance of the pathway

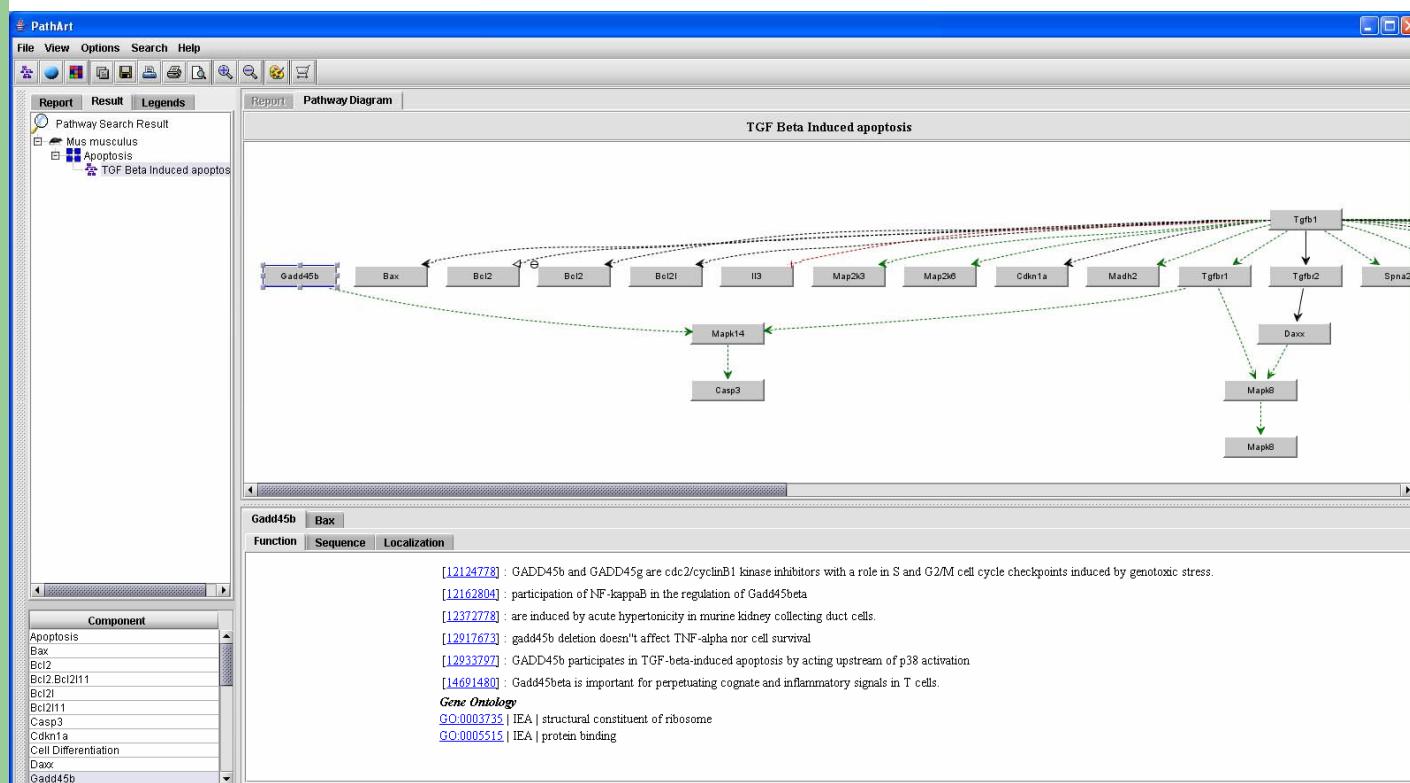
Component n...	Pathway Name	Organism	Physiology/...	IS_PHYSIO...	Pathway Ty...	Fisher P V...
Mapk14	CCR5 Mediated Pathway	Mus muscu...	Acquired I...	Disease	Signaling	0.089682
Gclc	Tat Signaling Pathway	Mus muscu...	Acquired I...	Disease	Signaling	0.236507
Mapk14	AGEs Signaling Pathway	Mus muscu...	Alzheimers	Disease	Signaling	0.149045
Tnfrsf1a	Amyloidbeta-peptide Signal...	Mus muscu...	Alzheimers	Disease	Signaling	0.118019
Tnfrsf1a	IFN Signaling Pathway	Mus muscu...	Alzheimers	Disease	Signaling	0.634464
Mapk14	LPS Signaling Pathway	Mus muscu...	Alzheimers	Disease	Signaling	0.386674
Map3k1	DR4/5 Signaling Pathway	Mus muscu...	Apoptosis	Physiology	Signaling	0.064863
Mapk14	IGF Signaling Pathway	Mus muscu...	Apoptosis	Physiology	Signaling	0.717471
Ngfr	NGF Signaling Pathway	Mus muscu...	Apoptosis	Physiology	Signaling	0.204750
Mef2d	T-cell apoptosis	Mus muscu...	Apoptosis	Physiology	Signaling	0.149045
Mapk14	TGF Beta Induced apoptosis	Mus muscu...	Apoptosis	Physiology	Signaling	0.051402
Tnfrsf1a	TNF Signaling Pathway	Mus muscu...	Apoptosis	Physiology	Signaling	0.776034
Mapk14	TNFR1 Signaling Pathway	Mus muscu...	Apoptosis	Physiology	Signaling	0.246824
Adprt	TRAIL Mediated Apoptosis	Mus muscu...	Apoptosis	Physiology	Signaling	0.378170
Mapk14	UV induced Antia apoptotic P...	Mus muscu...	Apoptosis	Physiology	Signaling	0.113880
Cd44	WNT Signaling Pathway	Mus muscu...	Apoptosis	Physiology	Signaling	0.644709
Mapk14	Interleukin Signaling Pathw...	Mus muscu...	Atheroscle...	Disease	Signaling	0.926753
Met	HGF Signaling Pathway	Mus muscu...	Breast Ca...	Disease	Signaling	0.064863
Tgfb2	TGF Signaling Pathway	Mus muscu...	Breast Ca...	Disease	Signaling	0.267067
Cd44	CD44 Signaling Pathway	Mus muscu...	Cell Adhe...	Physiology	Signaling	0.052216
Tp53	ATR-ATM Signaling Pathway	Mus muscu...	Cell Cycle	Physiology	Signaling	0.204750
Map3k1	Bcr-Abl Signaling Pathway	Mus muscu...	Chronic M...	Disease	Signaling	0.217718

Total genes = 60, Found genes= 17, Found pathways= 58





Interpretation (continued) PathArt





Interpretation (continued) GOFFA

- GOFFA – Gene Ontology For Functional Analysis
- Developed based on Gene Ontology(GO) database
- Grouping the genes into functional classes
- GO- three ontologies
 - Molecular function:** activities performed by individual gene products at the molecular level, such as catalytic activity, transporter activity, binding.
 - Biological process:** broad biological goals accomplished by ordered assemblies of molecular functions, such as cell growth, signal transduction, metabolism.
 - Cellular component:** the place in the cell where a gene product is found, such as nucleus, ribosome, proteasome.





Interpretation (continued) GOFFA

Select array type

Affy_RT-U34

Input gene list

Total original submit =15, Found =15 with GO term

No	LocusID	Gene Name...
1	24471	Hspb1
2	24472	Hspa1a
3	24553	Met
4	24561	Mpg
5	24565	Abcc1
6	24567	Mt1a
7	24605	Nras
8	24646	Abcb1
9	24811	Tap1
10	24842	Tp53
11	24862	Ugt2b
12	24888	Bcl2l1
13	24891	Abcb4
14	25035	Cyb5r3
15	25283	Glc



GOFFA

Tree Terms Genes GO Path GO Tree Prune

molecular_function

- all(15/553 P=1.000000 E=1.00 R=0.00)
 - molecular_function(15/522 P=0.416096 E=1.06 R=0.38)
 - enzyme regulator activity(1/25 P=0.504942 E=1.47 R=0.68)
 - caspase regulator activity(1/3 P=0.079327 E=12.29 R=1.78)
 - enzyme inhibitor activity(1/8 P=0.198624 E=4.61 R=1.38)
 - caspase inhibitor activity(1/2 P=0.053562 E=18.43 R=2.65)
 - transcription regulator activity(1/43 P=0.707856 E=0.86 R=0.53)
 - catalytic activity(10/345 P=0.478234 E=1.07 R=0.70)
 - signal transducer activity(2/104 P=0.807340 E=0.71 R=0.47)
 - transporter activity(3/43 P=0.103226 E=2.57 R=1.37)
 - binding(12/380 P=0.257422 E=1.16 R=0.97)
 - biological_process(13/486 P=0.730858 E=0.99 R=0.14)
 - regulation of biological process(6/179 P=0.349511 E=1.24 R=0.59)
 - response to stimulus(9/206 P=0.059628 E=1.81 R=1.36)
 - development(2/137 P=0.919836 E=0.54 R=0.17)
 - physiological process(13/464 P=0.555738 E=1.03 R=0.39)
 - cellular process(13/460 P=0.524454 E=1.04 R=0.42)
 - growth(1/19 P=0.412168 E=1.94 R=0.52)
 - cellular_component(14/457 P=0.233694 E=1.13 R=0.63)
 - Tree Terms Genes GO Path GO Tree Prune

GO path

Search

log₁₀(P) Value

Level

Top1 Top2 Top3 Top4 Top5 Top6 Top7 Top8 Top9 Top10

Tree Terms Genes GO Path GO Tree Prune

Molecular function Biological process Cellular component

No	Gene	Term	GO ID	Level (Aver.)	P value...	Gene Hits	E value
1	↑ <i>Abcb1</i>	nucleoside-triphosphatase activity	GO:0017111	7.00	0.001607	4.00	7.02
2	↑ <i>Abcc1</i>	nucleoside-triphosphatase activity	GO:0017111	7.00	0.001607	4.00	7.02
3	↑ <i>Nras</i>	nucleoside-triphosphatase activity	GO:0017111	7.00	0.001607	4.00	7.02
4	↑ <i>Tap1</i>	nucleoside-triphosphatase activity	GO:0017111	7.00	0.001607	4.00	7.02
5	↑ <i>Abcb1</i>	hydrolase activity, acting on acid anhydri...	GO:0016818	5.00	0.001933	4.00	6.70
6	↑ <i>Abcc1</i>	hydrolase activity, acting on acid anhydri...	GO:0016818	5.00	0.001933	4.00	6.70
7	↑ <i>Nras</i>	hydrolase activity, acting on acid anhydri...	GO:0016818	5.00	0.001933	4.00	6.70
8	↑ <i>Tap1</i>	hydrolase activity, acting on acid anhydri...	GO:0016818	5.00	0.001933	4.00	6.70
9	↑ <i>Abcb1</i>	pyrophosphatase activity	GO:0016462	6.00	0.001933	4.00	6.70
10	↑ <i>Abcc1</i>	pyrophosphatase activity	GO:0016462	6.00	0.001933	4.00	6.70
11	↑ <i>Nras</i>	pyrophosphatase activity	GO:0016462	6.00	0.001933	4.00	6.70
12	↑ <i>Tap1</i>	pyrophosphatase activity	GO:0016462	6.00	0.001933	4.00	6.70
13	↑ <i>Abcb1</i>	Tree Terms Genes GO Path GO Tree Prune	GO:0016462	6.00	0.001933	4.00	6.70
14	↑ <i>Abcc1</i>	molecular_function biological_process cellular_component	GO:0016462	6.00	0.001933	4.00	6.70
15	↑ <i>Nras</i>	GO tree prune	GO:0016462	6.00	0.001933	4.00	6.70

GO tree prune

Prune Tree

Pex: 0.05 Node size >= 5 E Value >= 2 Prune Tree Reset Tree Clear labels



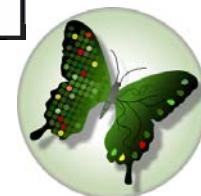
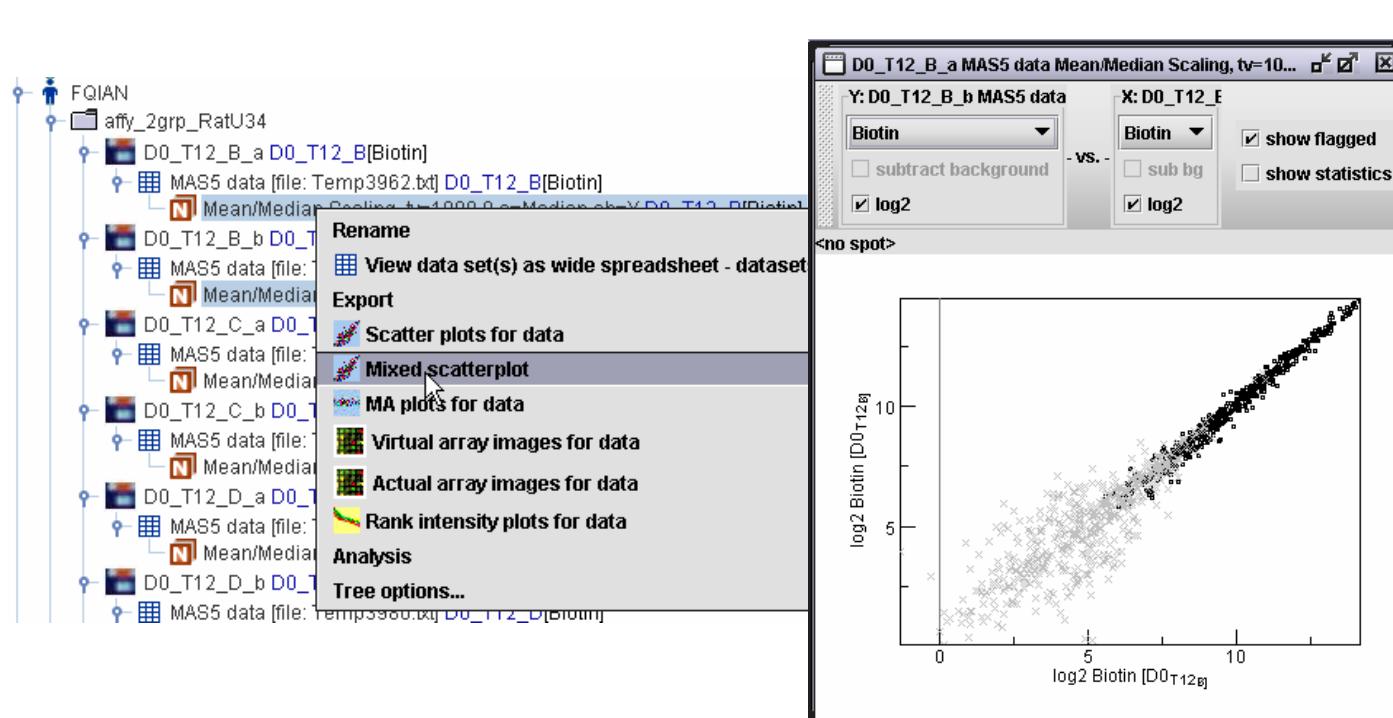
Data Exploring

- **Scatter Plot**
plot the fluorescence intensity data of Cy3 vs Cy5 for the same array
- **Mixed Scatter Plot**
compare two arrays in one plot, applies to both 2-channel and 1-channel data.
- **Correlation Matrix**
Correlation Matrix shows the correlation between column i and column j of the original matrix. It visually shows the correlation between two groups of data.
- **Bar Chart**
displays expression data for a single gene across multiple arrays within the same experiment or across different experiment.
- **Principal Component Analysis (PCA)**
PCA maps multidimensional data into a low dimensional graph (usually two or three dimensions) to visually inspect the sample relationships.
- **P-value Plot**
visual interpretation of P-value distribution.
- **HCA**
A two-way HCA is applied to investigate the grouping of samples in terms of their similarities in gene expression profiles, as well as the grouping of genes in terms of their similarities of samples.



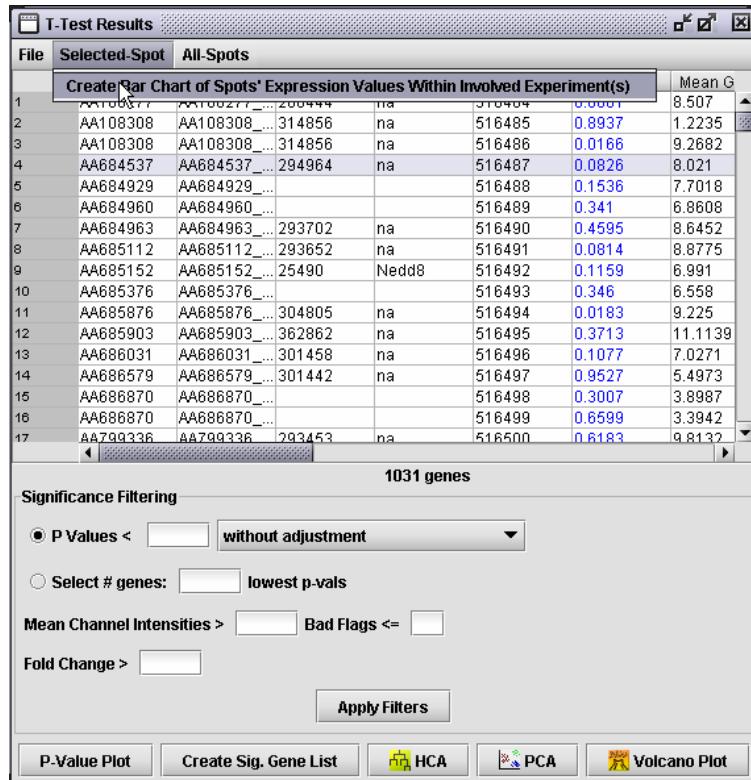


Data Exploring scatter plot





Data Exploring Bar chart



Access Bar chart from T-Test results.

Bar chart – display expression data for a single gene across multiple arrays in the same experiment or across different experiments.

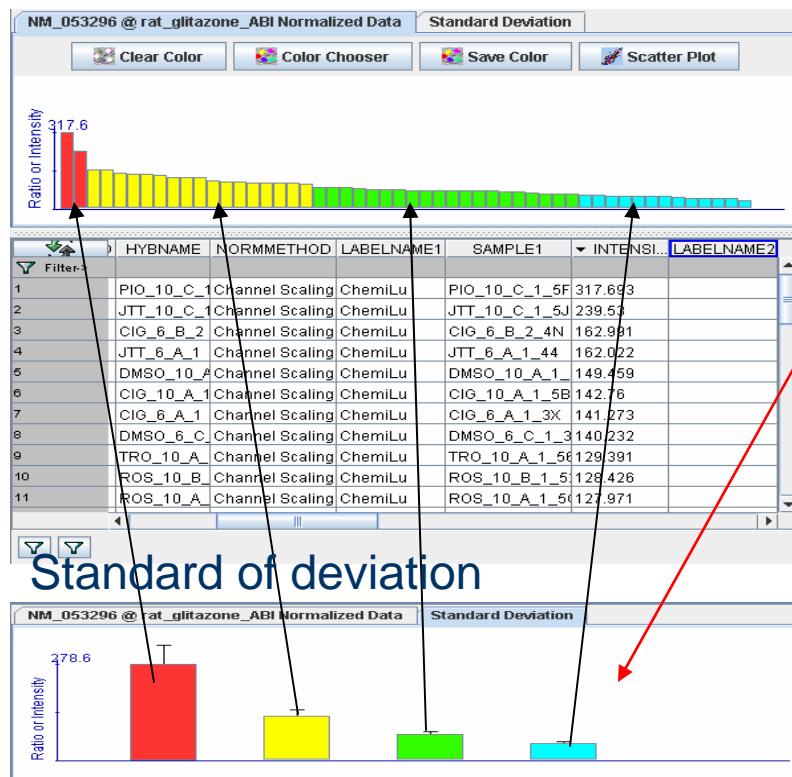




Data Exploring

Bar chart continued

Bar chart



Grouping multiple arrays across the same gene in different colors.

Group color could be saved.

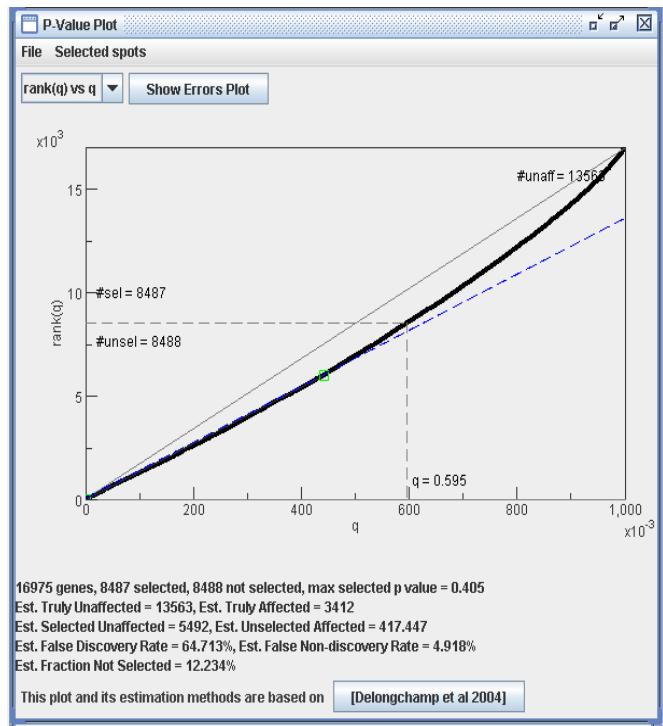
Standard deviation bar chart for the above groups. The bar height represents the mean intensity, while the T-line above the bar stands for the value of SD. The color for each bar echoes the color of the bar chart at the top, in the same order.



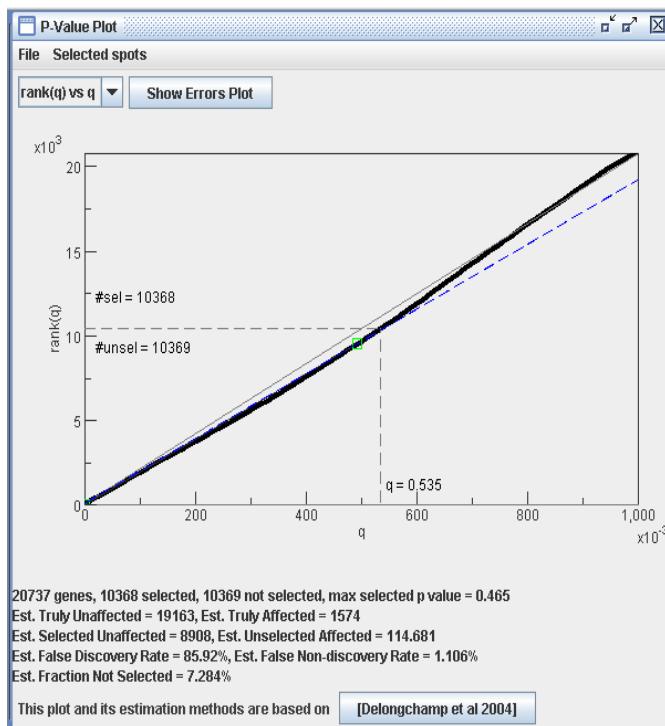
Data Exploring

P-value plot

Some effect

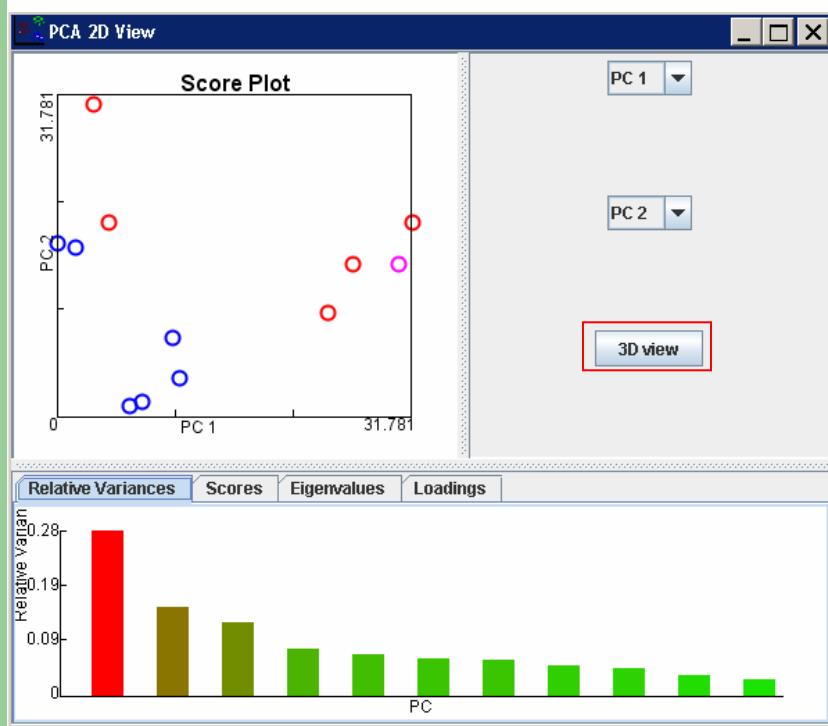


Slightly or no effect

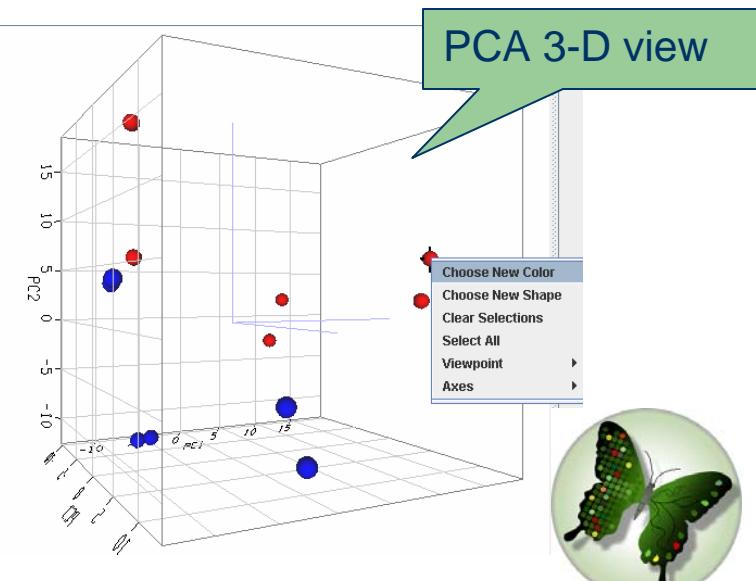




Data Exploring PCA

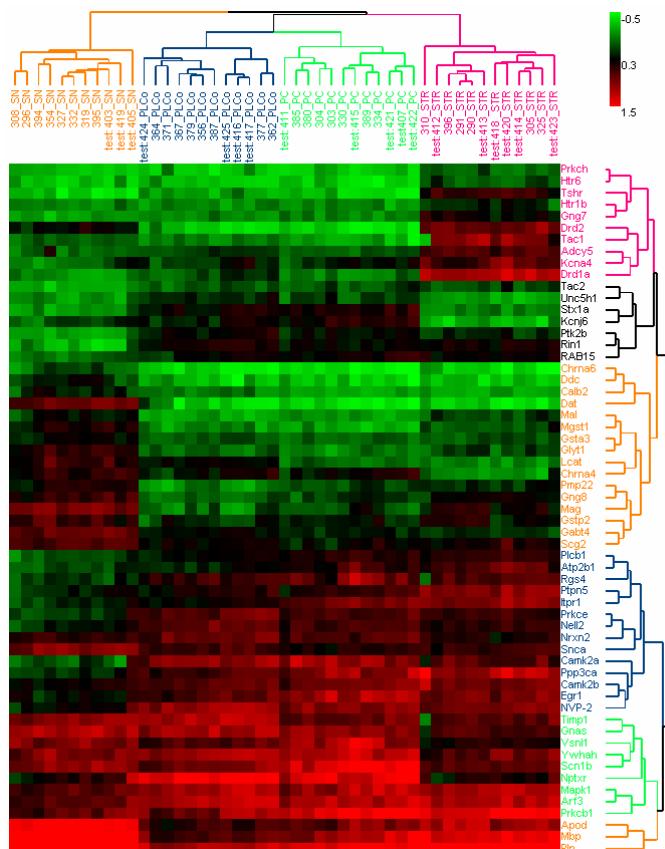


- Commonly used before gene selection
- To investigate the inter-sample relationship based on the gene expression profile
- Identify the outliers in biological/technical replicates
- View the variance of a multidimensional data





Data Exploring HCA



HCA is commonly used after gene selection for viewing the gene expression profile.





Accessing ArrayTrack

- FDA Internal:
<http://weblaunch.nctr.fda.gov/jnlp/arraytrack/index.html>
<http://weblaunch.nctr.fda.gov/jnlp/arraytrack/citrix/index.html>
- FDA External:
<http://edkb.fda.gov/webstart/arraytrack/index.html>





Technical Support

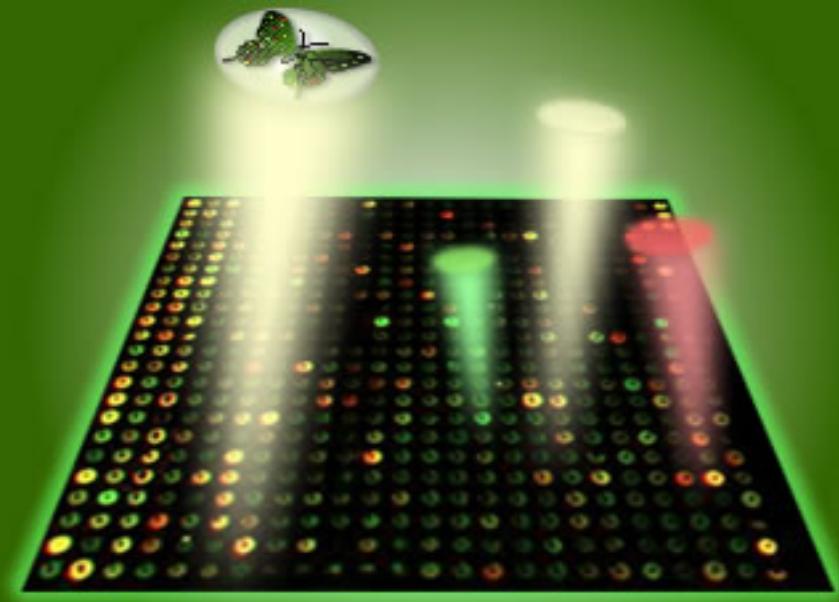
NCTRBioinformaticsSupport@nctr.fda.gov

ArrayTrack is developed by the U.S. Food and Drug Administration, National Center for Toxicological Research (FDA/NCTR).

FDA/NCTR reserves all rights for the software .



Thank you!



National Center for Toxicological Research
U.S. Food and Drug Administration